Published in final edited form as:

Res Dev Disabil. 2017 February ; 61: 32–43. doi:10.1016/j.ridd.2016.12.013.

“Didn’t I just say that?” Comparing parent report and spontaneous speech as indicators of grammatical development

Manya Jyotishi^a, Deborah A. Fein^a, and Letitia R. Naigles^a

^aDepartment of Psychology Sciences, University of Connecticut, 406 Babbidge Road, Unit 1020, Storrs, CT 06269

Abstract

Background—The Vineland Adaptive Behavior Scales use parental report to examine communication and social skills in children with typical and atypical development, and exhibit high reliability when compared against overall direct testing. However, findings are mixed comparing Vineland communication/language scores with experimenter-administered tests of language.

Methods—The current study breaks new ground in comparing Vineland reports with direct observation of children’s speech by (a) individual items and (b) level of child functioning, focusing on usage of wh-questions, verb tenses, negation, pronouns and noun-verb combinations. Both ‘high-verbal’ (HV) and ‘middle-verbal’ (MV) children with ASD are included, as well as a language-matched TD group.

Results—The results revealed that parent report on the Vineland varies in accuracy of capturing the production of grammatical items by young children with ASD and TD children. While parents’ assessment of their child’s production of noun-verb combinations and ‘who/why’ was highly accurate, children’s production of pronouns was under-rated by parents. Additionally, parents of HV children also under-rated their child’s production of past regular verbs.

Conclusion—Underestimation of these grammatical elements could lead to mistaken conclusions about their development in ASD or in individual children.

Keywords

Vineland; ASD; spontaneous speech; language assessment; grammatical items

1. Introduction

Researchers and clinicians interested in the language of children with Autism Spectrum Disorder (ASD) rely on both direct testing of children’s skills and parent report. Direct

Corresponding author: Manya Jyotishi, Department of Psychology, University of Connecticut, 406 Babbidge Road, Unit 1020, Storrs, CT, 06269 – 1020, USA. manya.jyotishi@uconn.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

evidence of children's language production and/or comprehension is frequently desirable so as to enable detailed analyses of specific construction and/or word usage. Such analyses inform both researchers interested in the language competence of children with ASD and clinicians concerned with their interventions. However, parent report about children's language may be desirable in a number of situations; for example, some children, particularly very young children or those who have not yet had any intervention, may lack the ability to sustain attention in a testing situation or may have decreased motivation to comply with adult requests (Koegel, Koegel & Smith, 1997). Parent report about their children's receptive and expressive competence may need to be relied on in such situations. Furthermore, some language behaviors are more likely to be demonstrated in social settings such as school or home, rather than in a laboratory or clinic; thus, parent report may be the only way to document these. Researchers and clinicians thus usually choose the language assessments that yield the most data relevant to their questions. Multiple studies have demonstrated a reasonable degree of agreement between total receptive and expressive scores as measured by direct assessment and parent report (e.g., Dale, 1991; Luyster, Kadlec, Carter & Tager-Flusberg, 2008); however, what is less often studied is how consistent are these different modes of assessment at the level of specific aspects of language. In the current study, we examined same-day agreement between parent report and child speech, for seven grammatical constructions.

A number of studies have documented that parental assessments of their child's cognitive, reading, and/or language ability are generally reliable with experimenter (direct) assessments for both typically developing (TD) and developmentally disordered children. For example, parental assessments yield similar rankings of the children in terms of ability. Oliver et al. (2002) compared TD 3-year-olds' performance on the McCarthy Scales (McCarthy, 1972) with their parents' at-home assessment of memory, perception, verbal, and quantitative tasks, and found that the McCarthy Scales and parent report correlated robustly and significantly ($r_s = 0.31 - 0.49$) (see also Glaun, Cole, Reddinough, 1999). Similarly, Feldman et al. (2005) compared parental report on the MacArthur Communicative Development Inventory (MCDI; Fenson et al., 1993) with TD 3-year-olds' performance on the McCarthy Scales and the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 1981), and found that all three MCDI measures (vocabulary, sentence complexity, using language) correlated significantly with the McCarthy ($r_s = 0.47 - 0.49$) and PPVT ($r_s = 0.41 - 0.49$). Dale (1991) found concurrent validity for parent report on the MCDI against standardized testing (e.g., Expressive One-Word Picture Vocabulary Test, $r_s = 0.34 - 0.73$) and a language sample (to judge MLU, $r_s = 0.33 - 0.76$), with 24-month old children. With school age children, too, parent observations of their TD children's language and reading correlated significantly with experimenter administrations of the Clinical Evaluation of Language Fundamentals (CELF; Semel, Wiig, & Secord, 1995; CELF Receptive: $r = 0.55$, CELF Expressive: $r = 0.60$) with those same children (Massa, Gomes, Tartter, Wolfson, & Halperin, 2008).

Furthermore, with children with developmental disorders such as Autism Spectrum Disorder (ASD) and Fragile X Syndrome (FXS), parental assessments of language and reading using the Vineland Adaptive Behavior Scales (Sparrow, Cicchetti & Balla, 2005) have been found to yield high levels of agreement with experimenter administered tests such as the Mullen

Scales of Early Learning (Mullen, 1995, Mullen Receptive: $r = 0.53$, Mullen Expressive: $r = 0.85$), the Neale Analysis of Reading Ability (NARA; Neale, 2007, Passage-level accuracy: $r = 0.81$, Passage-level comprehension: $r = 0.88$), the Early Language Milestones Scales (Coplan, 1993, $r_s = 0.24 - 0.50$), the Receptive-Expressive Emergent Language Scale (Bzoch & League, 1991, $r_s = 0.47 - 0.67$), and the Wide Range Achievement Test (WRAT; Wilkinson & Robertson, 2006, $r = 0.70$) (Luyster et al., 2008; Arciuli, Stevens, Trembath, & Simpson, 2013; Kover, McCary, Ingram, Hatton, & Roberts, 2015). Additionally, researchers have demonstrated that parent reports can yield similar categorizations of developmental delay. For example, Johnson, Wolke and Marlow (2008) found high sensitivity and specificity between their parent report instrument and the Bayley Scales of Infant Development (Bayley, 2006), with their identifications of developmental delay in 2-year-olds born premature. This agreement has some limits, though. For example, Feldman et al. (2005) reported that parent assessments of 2-year-olds' vocabulary on the MCDI were only 'fair to good' predictors of children's categorization as clinically language delayed at the age of 3; moreover, while Duff, Reen, Plunkett and Nation (2015) found that vocabulary size by parental report during infancy predicted vocabulary and reading performance during school age, infant vocabulary accounted for only a modest percent of the variance (under 20%).¹

There are also some indications that parent report is not a reliable substitute for all measures of child development. For example, Feldman et al. (2005) found that parent assessments of their two-year-olds' grammatical levels (i.e., sentence complexity on the MCDI-WS), did not correlate with any other concurrent or longitudinal language measures. Furthermore, parent estimates of the developmental ages of their children with ASD seem to be consistently overestimated (Geiger, Smith & Creaghead, 2002). Interestingly, even though children's Vineland and Mullen scores are reported to be correlated, Vineland reports of *expressive* language in toddlers with ASD have yielded consistently lower age-equivalent scores than the experimenter-administered Mullen whereas Vineland reports of *receptive* language have yielded consistently higher age-equivalent scores relative to the Mullen (Luyster et al., 2008). Furthermore, correlations between parent report and direct testing tend to be higher for expressive than receptive language (Luyster et al, 2008; Ring and Fenson, 2000). Thus, parents may under-estimate their child's expressive language levels and/or overestimate their comprehension abilities; the latter finding may be due at least in part to parents being unaware of the non-linguistic cues they are providing (Luyster et al, 2008). More generally, Geiger et al. (2002) observed that parent estimates of their children's levels of functioning were more disparate if the children were lower-functioning, and Ozonoff et al. (2011) found that parent report via the ADI-R of their children with ASD's early developmental trajectory was frequently inconsistent with data provided by the home videos that parents themselves recorded. That is, the home videos differentiated children's early developmental trajectories into three categories (early onset, regression, and plateau); however, parent report frequently missed the regression pattern. Such inconsistencies between parent report and standardized-test or behavioral/expressive language measures warrant a closer look, which we undertake in the current study.

¹Of course, the low levels of longitudinal predictability need not be the result of parental report measures in particular.

Another concern about the reliability of parent report involves item-specificity. That is, parent report measures yield scores calculated across items, and experimenter-administered measures yield scores calculated across items, and the consistency or reliability of the two types of measures has always been calculated based on these overall scores, rather than based on the reliability or agreement of specific items. For example, vocabulary assessments are usually judged for reliability based on overall words produced or understood on vocabulary checklists, on experimenter assessments, and in children's speech (Fenson et al., 1993; Kover et al., 2015; Libertus, Odic, Feigenson, & Halberda, 2015; Dale, 1991). Grammatical assessments are usually judged for reliability based on total number of grammatical forms that parents endorse, total scores on experimenter assessments, and MLU in children's speech (Fenson et al., 1993; Feldman et al., 2005; Dale, 1991). With respect to vocabulary, recent research has begun to compare children's performance on preferential looking tasks (i.e., looking longer to named items over foils) with the words their parents say they produce or understand (e.g., Styles & Plunkett, 2009). In contrast, little research has examined whether parent report of children's use of specific grammatical items, such as pronouns or the past tense or wh-words, agrees with experimenter assessments of those same grammatical forms, or even of children's productions of those forms in their actual spontaneous speech.

Why should parental endorsement of the presence or absence of specific grammatical forms matter? There is some indication that specific grammatical forms are of particular difficulty for children with developmental disorders. For example, pronouns, wh-questions, and tense markers are frequently mentioned as key linguistic difficulties for children with ASDs (Eigsti, Bennetto & Dadlani, 2007; Park, Yelland, Taffee & Gray, 2012; Tager-Flusberg, Paul & Lord, 2005), especially those with co-morbid grammatical impairments (Norbury, 2013). The degree to which these difficulties can be identified primarily from parent report (i.e., without a speech sample), though, is unknown. In addition, there has been no investigation of how the reliability of parent report of specific grammatical items might vary depending on the child's level of functioning and/or developmental trajectory. As mentioned above, some children are difficult to test because of attention or compliance issues; if they do not produce a grammatical form in a language sample or on standardized testing in a strange environment, they may still be competent in that area. Parent report of the skills they display in other environments may be the only measure available. The purpose of the current study was to conduct a detailed comparison of children's usage of seven grammatical constructions—pronouns, negation, noun-verb combinations, past tense, progressive aspect, what/where questions, and who/why questions—in their spontaneous speech and as endorsed by parent report. As part of a longitudinal project, we collected children's speech during parent-child play sessions from 20 children with ASD at six visits, four months apart. We also collected parent report via the Vineland Adaptive Behavior Scales (Vineland; Sparrow, Cicchetti & Balla, 2005) at each visit. The children varied widely in language ability; therefore, we were able to compare children whose language levels were age-appropriate with children who were verbal but language-impaired. We also explored the degree to which the parent report and spontaneous speech of TD children, at similar language levels, were reliable on these grammatical items.

2. Method

2.1. Participants

Twenty children with an ASD (mean age at visit 1 = 32.12 months, SD = 5.73) and twenty-four TD children (mean age at visit 1 = 20.40 months, SD = 1.67) participated in this longitudinal study. In the ASD group, the participants included three girls and 17 boys; these children were recruited through treatment facilities in the Northeastern United States and diagnosed with Autistic Disorder or Pervasive Developmental Disorder – Not Otherwise Specified (PDD-NOS) by professionals. Children in the TD group were recruited using birth announcements from newspapers. Participants in this group included three girls and 21 boys. Family income levels in both groups ranged from lower-SES to upper-SES. Diagnostic status was verified at visit 1 via the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 1989; see Table 1). Children with ASD who did not meet the diagnostic criteria when the ADOS was administered were not included in the study. In the ASD group, there were missing data due to audio problems at visit 5 for one male participant, and at visits 3 and 6, respectively, for another two female participants. There were no missing data for participants in the TD group.

Because this study was designed to compare children by language levels, we matched the TD and ASD groups at visit 1 on expressive language (i.e., the “Total Understands and Says” section of the MCDI; Fenson et al. 1993; $t(42) = -0.59, p = 0.56$). We matched the groups on the conservative criterion of a p-value of 0.50 for the group difference (Mervis & Klein-Tasman 2004). An additional 13 children with ASD participated in the overall study (Naigles & Chin, 2015; Naigles & Fein, 2017); however, these children were not included in the current study because they remained nonverbal (i.e., produced none of the critical items on the Vineland Communication subscale) throughout the six visits.

Based on the findings of Tek, Mesite, Fein and Naigles (2014), who reported two distinct developmental trajectories of word and grammatical morpheme use in our overall ASD sample, we further subdivided the ASD group in the current study into high-verbal (HV) and middle-verbal (MV) subgroups, according to their Mullen Expressive Language T-Scores at visit 6. Children whose expressive score was within the normal range (40+) were sorted into the HV-ASD group, and children whose expressive score was ≤ 1.0 SD normal range (lower than 40) were sorted into the MV-ASD group. There were 9 children in the HV-ASD group and 11 children in the MV-ASD group. Note that, as shown in Table 1 and as previously reported by Tek et al. (2014), the HV-ASD group, being chronologically older, had higher expressive language levels than the TD group at Visit 1. That is, the whole ASD group was matched with the TD group at Visit 1 but after the exclusion of nonverbal children, the two groups were no longer matched.

2.2. Procedure

The participants' data were collected across six home visits, each of which was separated by 4 months. During the visits, children were administered standardized tests; then, they and their parent engaged in a 30-minute semi-structured parent-child play session. At the end of

each visit, the parent who had engaged in the play session (almost always the mother) was administered the Vineland in interview format.

Play sessions were videorecorded, and the children's speech measures were derived from this session. The first 15 minutes of the session followed the structure of the Screening Tool for Autism in 2-year-olds (STAT, Stone, Coonrod, & Ousley, 2000), which consists of 12 play-based activities that involve the child in pretend play with dolls, interactive play with a ball or truck, imitative action play, and requests and joint attention (e.g., pointing, reaching, etc.). During the structured play session, parents were handed instruction cards that prompted them to play with their child with items provided by the researcher. Thus, at least half of each session followed the same protocol and could enable the same type(s) of grammatical structures (e.g., wh-questions, pronouns, etc.). During the second (free play) part of the session, the parent and the child were instructed to play "however they usually play at home." Recent analyses of these types of play sessions have indicated that they are successful at eliciting a large number of child utterances (Kover, Davidson, Sindberg & Weismer, 2014).

2.3. Standardized Test Measures

The first three measures were collected to confirm the children's placement into diagnostic groups, and to provide general characteristics of their language levels. All standardized tests were administered by the experimenter.

The *Autism Diagnostic Observation Schedule* (Lord et al., 1989) is a structured and play-based assessment for the diagnosis of autism spectrum disorders. It consists of a series of activities designed to interest young children and encourage them to communicate, and systematic probes are used to sample children's behavior in social interaction, communication, stereotypical behavior and repetitive interests. The ADOS was administered at visits 1 and 5.

The *Mullen Scales of Early Learning* (MSEL; Mullen 1995) is a measure of intellectual development, which includes items that measure visual reception, expressive and receptive language, and motor development for children from birth to 5 years, 8 months. The MSEL gives raw scores, *T* scores (average *T*-score is 50 with a standard deviation of 10), and age equivalents for each domain of the test. The MSEL was administered at visits 1 and 6.

The *MacArthur Communicative Development Inventory* (Fenson et al. 1993) is a standardized parent reporting instrument used to assess early language development. The infant version (Words & Gestures) was administered at Visit 1. We report here the results of the vocabulary checklist, measuring children's production of a possible 396 words.

The *Vineland Adaptive Behavior Scales, 2nd edition* (Vineland II; Sparrow et al., 2005) is a parent report measure that assesses a child's communication, daily living skills, socialization and motor skills; it is administered to a parent as a semi-structured interview. Parents' responses are rated on each question using a scale of 0 (never present), 1 (sometimes present), 2 (always present) or DK (don't know). The Vineland yields both subdomain standard scores and an overall standard score (see Table 1). For the purposes of this study,

we further analyzed seven items, which targeted early use of grammar, in the Expressive-Communication subscale, repeated verbatim below (items assessing prepositions were not included because the play sessions did not consistently afford numerous opportunities for preposition usage).

18. Noun-verb combinations: Uses phrases with a noun and a verb (for example, “Katie say”; “Go home”; etc.).

22. What/Where questions: Asks questions beginning with *what* or *where* (for example, “What’s that?”; “Where doggie go?”; etc.).

23. Negations: Uses negatives in sentences (for example, “Me no go”; “I won’t drink it”; etc.); grammar is not important.

31. Who/Why questions: Asks questions beginning with *who* or *why* (for example, “Who’s that?”; “Why do I have to go?”; etc)

32. Progressive verbs: Uses present tense verbs ending in *ing* (for example, “Is singing”; “Is playing”; etc.).

34. Pronouns: Uses pronouns in phrases or sentences; must use correct gender and form of the pronoun, but sentences need not be grammatically correct (for example, “He done it”; “They went”; etc.).

36. Past tense verbs: Uses regular past tense verbs (for example, walked, baked, etc.); may use irregular past tense verbs ungrammatically (for example, “I runned away”; etc.).

Maternal responses to these items were coded as ‘says’ if mothers had marked (1) or (2), and ‘doesn’t say’ if mothers had marked (0) or DK.

2.4. Spontaneous Speech Coding

Spontaneous speech during the play sessions was transcribed using CHAT by two transcribers. Ten percent of the transcripts were independently transcribed for reliability. The transcripts were then compared and all differences were resolved by discussion and review of the recordings; therefore, reliability of the final transcript was very high, $r = .99$. The transcripts were coded using CLAN (MacWhinney, 1995); every instance of the seven Vineland items was extracted and tabulated from each transcript. The transcript length, on average, consisted of 147.45 child utterances ($SD = 73.32$) at the early visits (visits 1–3). During the early visits, TD children produced 144.94 utterances ($SD = 73.21$), children with HV-ASD produced 196.37 utterances ($SD = 70.35$) and children with MV-ASD produced 101.03 utterances ($SD = 76.41$). During the later visits (visits 4–6), the transcript length consisted of 201.64 child utterances ($SD = 80.38$); TD children produced 206.92 utterances ($SD = 82.65$), children with HV-ASD produced 238.56 utterances ($SD = 76.15$) and children with MV-ASD produced 159.45 utterances ($SD = 82.35$).

At each visit, each child’s rating (‘says’ (code of 1 or 2), ‘doesn’t say’ (code of 0)) on each Vineland item was compared with his/her spontaneous speech. Two codes were of interest.

(1) For each Vineland item in the transcript (i.e., produced during the interactions), we coded whether the mother confirmed this on the concurrent Vineland or not (accurately confirmed/not confirmed). (2) For each Vineland item reported by the mother as ‘doesn’t say’ (i.e., 0), we coded whether this linguistic form was in the concurrent transcript (Inaccurately overlooked/correct). We did not further analyze those items that were unconfirmed in the transcripts, because the absence of an item in the transcript does not necessarily mean the child does not say this item (i.e., he or she may say it in a different context).

2.5. Analyses

To capture developmental changes that have previously been observed with this dataset (Tek et al., 2014; see also Naigles et al., 2016), we combined the data from visits 1–3 into ‘early visits’ and those from visits 4–6 into ‘later visits’. The goals of the analyses were twofold: First, to ascertain the overall degree of consistency between maternal report and child speech, and second, to compare the grammatical forms to discover if some are more vulnerable to maternal report-child speech mismatches.

For each group (TD, HV-ASD, MV-ASD), and within the Early and Later visits, we first summed the number of instances of accurately confirmed (present in transcript + ‘says’), and of inaccurately overlooked (‘doesn’t say’ + present in transcript) for each item across all transcripts and Vineland administrations. For example, with 24 children in the TD group, there were 72 transcripts and 72 Vineland administrations across the first three visits. Possible scores for a given item for accurately confirmed ranged from 72 (each child produced that item at each visit, and each mother reported ‘says’ on each Vineland administration) to 0 (some children produced that item at some visits, but no mother reported ‘says’ on any Vineland administration). Possible scores for a given item for inaccurately overlooked also ranged from 72 (all mothers reported ‘doesn’t say’ at each Vineland administration and all children produced that item at each visit) to 0 (some mothers reported ‘doesn’t say’ for some Vineland administrations and none of the children of those mothers produced that item at those same visits). Items for which no mothers in a group reported ‘doesn’t say’ (which occurred at some of the later visits) were excluded for that group. Then, the overall maternal percent accurately confirmed for a given item was calculated as the number of accurately confirmed instances (i.e., across all Vineland administrations) divided by the total number of items (i.e., across all transcripts). Overall maternal percent for inaccurately overlooked for a given item was calculated as the number of instances (i.e., across all transcripts) divided by the total number of instances for which mothers reported ‘doesn’t say’. Mean percent maternal accurately confirmed and mean percent maternal inaccurately overlooked were calculated by averaging across the seven items. These means were then compared across visits using the paired-sample Wilcoxon test, and across groups using the Mann-Whitney U test.

For the second analyses, we compared these mean percent accurately confirmed and percent inaccurately overlooked levels with the percent accurately confirmed/inaccurately overlooked for each item separately, using the binomial equation, to ascertain whether individual items followed the group pattern.

3. Results

3.1. Descriptives of child speech and maternal report

Table 2 presents the percent of transcripts demonstrating the children's productions of the seven Vineland items; these increased significantly from the early to later visits for each group ($Wobs = 0$, $p < .05$). Moreover, a greater proportion of transcripts from the HV-ASD group demonstrated the children's production of the items compared with transcripts from the MV-ASD group at both visits (Us ($N=14$) = 2 (early), 3 (later), $p < .005$), and compared with the TD group at the early visits (U ($N=14$) = 5, $p = .013$) but not the later ones ($p = .259$). A greater proportion of transcripts from the TD group demonstrated the children's productions of the items compared with transcripts from the MV-ASD group at both visits (Us ($N=14$) = 9 (early), 6 (later), $ps < .05$). Interestingly, the proportion of transcripts from the HV-ASD group that demonstrated children's productions of the items at the early visits did not differ from the proportion of transcripts from the MV-ASD group that demonstrated children's productions of the items at the later visits (U ($N=14$) = 11, $p > .05$); this is consistent with the observation that the HV-ASD group's Mullen EL and RL scores at Visit 1 did not differ significantly from the MV-ASD group's Mullen EL and RL scores at Visit 6 (see Table 1).

Table 3 presents the mean percent of maternal reports of 'doesn't say' for the seven Vineland items for the early and later visits. The total administrations for a given item were frequently fewer than the total number of administrations because, especially at the early visits, the mothers did not reach the later items once they had generated three 'doesn't say' responses in a row; hence, these Ns are smaller. For instance, there were 22 administrations in the TD group that were not counted at the early visits because the mothers did not reach some of the later items on the Vineland. For this reason, we include the number of administrations in each table. At the early visits, the TD and HV-ASD groups both reported smaller proportions of 'doesn't say' responses than the MV-ASD group (TD vs. MV-ASD: U ($N=14$) = 5, $p = .011$; HV-ASD vs. MV-ASD: U ($N=13$) = 6, $p = .035$); the TD and HV-ASD groups did not differ from each other. The mean percentage of children reported to not say the items decreased from the early to later visits in all three groups; statistical comparisons involving the TD and HV-ASD groups at the later visits were not performed because only a few items were applicable for these groups. However, the 'doesn't say' responses for the MV-ASD group did decrease significantly from the early to later visits ($Wobs = 0$, $p < .05$).

3.2. Overall levels of maternal confirmation and incorrectness

Our first set of Vineland-Speech comparison analyses examined the degree to which mothers accurately confirmed their children's speech, and children's speech supported their mothers' reports of "Does not say," across all seven Vineland items. Table 4 presents the percent of maternal accurate confirmations of children's productions at the early and later visits. Paired-sample Wilcoxon tests revealed that maternal accurate confirmation percentages were significantly higher at the later visits for all three groups ($Wobs = 0$ or 1, $ps < .05$). Non-parametric tests revealed no group differences in the distribution of maternal percent accurately confirmed at early visits; however, Mann-Whitney U tests yielded statistically significant group differences at the later visits between the TD and MV-ASD groups ($U =$

4.0, $N = 14$, $p = .007$) and between the HV-ASD and MV-ASD groups ($U = 3.5$, $N = 14$, $p = .004$), with maternal accurate confirmations being lower in the MV-ASD group in both comparisons. No significant differences were found between TD and HV-ASD children during either the early or later visits.

Table 5 presents the percent of maternal inaccurately overlooked structures. For all groups of children, mothers missed substantial numbers of their child's constructions; approximately one-third of all 'does not say' maternal responses at the early visits were *disconfirmed* by the transcripts. That is, mothers reported 'does not say' for a given item, but one-third of the time, children actually did produce that item at least once in the transcript at that visit. This level of maternal incorrectness did not differ by group at the early visits; group differences for the TD and HV-ASD groups cannot be assessed statistically at the later visits because of the small number of relevant items. For the MV-ASD group, maternal percent for inaccurately overlooked structures was significantly higher at the later than early visits ($Wobs = 0$, $N = 7$, $p < .05$).

3.3. Levels of maternal confirmation and incorrectness by item

The second set of Vineland-Speech comparison analyses investigated whether mothers were equivalently inaccurately overlooking the seven Vineland items. Figure 1 shows the percent inaccurately overlooked by each Vineland item for all three groups during the early and later visits, with each item designated whether it was incorrect at levels higher than, lower than, or not different from average, using the binomial equation (The averages across items are provided in Table 5).

For the TD children, the pronoun construction yielded the highest inaccurately overlooked rates, such that *every* child whose mother reported 'does not say' produced at least one pronoun correctly at the same visit. In contrast, mothers of TD children reported inaccurately overlooked structures at rates below the group average with 'who/why' questions at both early and later visits, and with noun-verb combinations at the early visits. For the HV-ASD group, maternal reports of 'does not say' for both the regular past tense on verbs, and for pronoun use, were inaccurate at higher-than-average levels at the early visits. For the MV-ASD group, maternal reports of 'does not say' inaccurately overlooked their children's speech at levels above the group average for pronouns at both early and later visits; they were inaccurate at below average levels for what/where questions, who/why questions, and production of the regular past tense.

4. Discussion

This study was conducted to investigate the degree to which parent report of children's production of seven grammatical items, on the Vineland, was consistent with those children's actual speech produced during a play session with the same parent on the same day. Vineland and speech assessments were collected at six visits, four months apart, and both corroborated the expected developmental pattern that more children produced these seven grammatical items at the later visits than the early ones, in all groups. Similarly, parental reports that their child 'doesn't say' these items were fewer in number at the later visits compared to the early visits. With regard to parental confirmations of children's

speech, mothers of TD and children with HV-ASD accurately confirmed child productions at generally high rates, but did improve (i.e., from 80% to almost 100%) from the early to later visits. Mothers of children with MV-ASD accurately confirmed significantly fewer of their children's productions at both visits (57% to 64%). With regard to the parental reports of 'doesn't say', approximately one-third of these were inaccurately overlooked at the early visits—parents reported that children did not say items that they did in fact produce at those visits. At the later visits, mothers of children with MV-ASD were even less accurate, with about half of their reports of 'doesn't say' disconfirmed by the children's transcripts. Pronouns were particularly challenging to catch: mothers reported that children did not say these—when they actually did—at rates higher than the group average in all three groups. Finally, mothers of children with HV-ASD frequently reported 'does not say' for the regular past tense, when their children produced these at the same visit.

Our findings shed light on three important issues concerning the use of parent report for assessing the use of specific grammatical items by young children with ASD. First, parent levels of accurate confirmation and inaccurate overlooking differed by children's language levels, with more items that had been produced in the transcripts being missed by parents of children with MV-ASD than by TD children or children with HV-ASD. It seems likely that mothers of the less fluent talkers (the MV-ASD group) may not be used to hearing those constructions that are just emerging in their children's speech, and thus they did not remember the productions when queried on the Vineland (Geiger et al., 2002). In contrast, mothers of the other two groups, which included children who were already fluent talkers, have evidently heard these constructions often enough to be able to 'picture' them when asked. Supporting this interpretation is our finding that the accurate confirmations and inaccurate overlooks of the mothers of the MV-ASD group at the later visits were similar to those of the mothers of the HV-ASD group at the earlier visits, consistent with the observation that the language scores of the HV-ASD group at visit 1 were similar to those of the MV-ASD group at visit 6. It seems likely, then, that the higher incidence of inaccurate overlooks and lower incidence of accurate confirmation, in the MV-ASD group, was attributable to their children's relative sparsity of speech. This tendency supports our suggestion that collecting language samples is a critical component to determining the grammatical development of children with ASD.

It is also possible that the general language delay of the children in the MV-ASD group led to lowered expectations and a negative halo effect. These parents might be focusing on their child's acquisition of nouns and verbs, which are often targeted in ABA programs. Furthermore, the children may have picked up the past tense and other grammatical markers initially through echoing the adult's questions or statements rather than through direct teaching, and parents might discount or pay less attention to the child's productions that were not specifically taught.

The second important issue concerns the overall relatively high level of inaccurate overlooks on the Vineland for these seven grammatical items. Again, approximately one-third of the time at the early visits, a child produced a grammatical construction at least once in a transcript/play session, yet the parent who participated in that play session subsequently stated that the child did not produce that item at all (i.e., endorsed a '0' on the Vineland).

This level of inaccuracy did not differ across groups at the early visits; therefore, it does not seem to be a characteristic of parents of children with ASD in particular. Instead, we suggest that this level of inaccuracy is a characteristic of how parents of young child language learners reflect on their children's language, especially their grammatical usage. While it seems that parents are generally accurate remembering the specific words that their young children say (Styles & Plunkett, 2009), remembering their specific grammatical constructions seems much harder. This is consistent with findings from the adult psycholinguistic literature, that adults are much more accurate at remembering the 'gist' of a given passage than the specific grammatical constructions used (e.g., Just & Carpenter, 1976). This level of inaccuracy should give researchers and clinicians pause, though, when considering how to use parent report measures; for example, parent report may not always be the most valid indicator of grammatical impairment in children with ASD, because parents may report their children not producing grammatical items such as the past tense when—at least some of the time—the children do produce these items. Any such inaccuracies would also lower the child's overall expressive language scores.

And indeed, the third important issue that our findings address was that the grammatical items differed in their vulnerability to parent report inaccuracy. In particular, parents in all three groups frequently missed pronouns and at both early and later visits; past tense markers were also missed at rates higher than the group average, for children in the HV-ASD group (for instance, nine mothers inaccurately reported 'doesn't say'). It is important to keep in mind that even though mothers of TD children at later visits apparently overlooked past tense markers a high *percentage* of the time, this was based on only *five* maternal responses of 'doesn't say', averaged across only *four* of the seven constructions. This percentage for the TD group should thus be interpreted with caution; therefore, we focus here on the HV-ASD group, whose group average is based on all seven constructions. Missing the past tense markers of the children in the HV-ASD group may be attributed to parental lower expectations; that is, some parents may have been told that their children were likely to experience difficulties producing the past tense, and so they were not listening closely for this marker. The fact that the past tense is often realized as a single segment [walk/t/, play/d/] can also make this marker more difficult to hear than, say, the progressive marker, which is a full syllable [walk-ing].

This 'salience' explanation may not extend to parental inaccuracy with pronouns, though, because pronouns are usually realized as full syllables in English, particularly in shorter sentences. Why did parents miss so many of their children's uses of personal pronouns? One consideration could be that the parents heard their children's pronoun uses, but did not understand the intent of the Vineland question. That is, parents may have remembered hearing that children with ASD frequently reverse pronouns (using 'I' for 'you'), may have heard their child reverse pronouns at least once and thus may have concluded that the child did not produce pronouns correctly at all. Alternatively, they may have heard their child use "I" and 'you' but not 'he' or 'they' and therefore concluded that the child did not produce all pronouns correctly. An interesting follow-up study to investigate the 'salience' vs. 'understanding' explanations could involve presenting videos that show children using the grammatical items with varying degrees of accuracy, and then asking parents whether they heard the children using the specific grammatical items in the video. Additionally, varying

the way these questions were asked would help us understand whether the parents truly overlooked the relevant utterances or did not understand the targets of the questions.²

In other analyses of this dataset, we have reported that these children did produce pronoun reversals, albeit at fairly low rates (i.e., averaging no more than 7% of all 1st and 2nd person singular pronoun uses; Naigles et al., 2016). For the current study, we revisited the transcripts to ascertain whether any child only produced pronoun reversals (or other pronoun errors) at a given visit. However, we found that every child who produced pronoun reversals or other pronoun errors at a given visit also produced as many if not more instances of correct pronoun usage (for more details about pronoun usage in this dataset, see Naigles et al., 2016; Keltz-Stephen, Tek, Fein & Naigles, 2014). Thus, these children should have been given a '1' for those visits; however, many were given a '0', presumably because parents were focusing on the noticeable errors, rather than the many correct uses. In fact, a general point could be made that in comparing parents' Vineland responses to examining transcripts, the latter gives credit to single instances of correct usage, while parents may respond according to what they see as the most typical behavior. Given this level of inaccuracy, though, parent lack of endorsement of the pronoun item should be interpreted with caution by researchers and clinicians.

4.1. Limitations

One limitation of the study is that we only included children with ASD who were verbal. Because these children are higher functioning, the findings may not generalize to all children with ASD even when and if they do start speaking. Also, the fact that the Vineland was not fully administered to all parents (i.e., parents stopped after 3 "0" responses) might have resulted in some subsequent items not being considered/analyzed for a given child. This led to lower *N*s because some items were not included for the groups, either because the children were all producing a construction at high rates, or not producing it at all. For example, if a parent reached the three "0" responses before the pronoun question, the use of pronouns for that child for that specific visit was not counted. Thus, only the questions that parents responded to for a given child were included. However, we do acknowledge that it would be beneficial to have larger *N*s. It would also be informative to further probe the parents about the extent of their explicit knowledge and ability to identify of these grammatical items; possibly, some items are more meta-linguistically available than others. Moreover, we did not examine the opposite error, when the child does not produce an item in the transcript but mother reports 'says' on the Vineland. This would require gathering a very large spontaneous language corpus to confirm these grammatical forms in order to determine whether the mother is right or wrong, so this remains an open question.

4.2. Conclusions

This small-scale study has found that parents of children with ASD are, overall, able to confirm their children's production of grammatical items on the Vineland, especially as the children become more verbal. Parents of lower verbal children are more likely to miss their children's grammatical productions, though, and the past tense marker and personal

²We thank one of our anonymous reviewers for suggesting such a follow up study.

pronouns are especially vulnerable to being missed. These findings suggest limits for interpreting parental reports of children's emerging grammar.

Acknowledgments

This research was funded by the National Institute on Deafness and Other Communication Disorders (NIH-DCD) [grant number R01 2DC007428]. We are grateful to Andrea Tovar for helping to launch this project, to Rose Jaffery and Janina Piotroski for their assistance in data collection, and to undergraduates at the UConn Child Language Lab for their assistance in coding. Portions of this research were presented at the International Meetings for Autism Research and the Biennial Meeting of the Society for Research in Child Development, and we are grateful for comments and suggestions received at those venues. Finally, we are extremely grateful to the families who have participated in this research.

References

- Arciuli J, Stevens K, Trembath D, Simpson IC. The relationship between parent report of adaptive behavior and direct assessment of reading ability in children with autism spectrum disorder. *Journal of Speech, Language, and Hearing Research*. 2013; 56(6):1837–1844.
- Bayley, N. Bayley Scales of Infant and Toddler Development. 3. San Antonio: TX: Harcourt Assessment; 2006.
- Bzoch, K., League, R. Receptive – Expressive Emergent Language Test: A method for assessing the language skills of infants, second edition. Austin, TX: Pro-Ed; 1991.
- Coplan, J. Early Language Milestone Scale – Second edition (ELM-2). Austin, TX: ProEd; 1993.
- Dale PS. The validity of a parent report measure of vocabulary and syntax at 24 months. *Journal of Speech, Language, and Hearing Research*. 1991; 34(3):565–571.
- Duff FJ, Reen G, Plunkett K, Nation K. Do infant vocabulary skills predict school – age language and literacy outcomes? *Journal of Child Psychology and Psychiatry*. 2015; 56(8):848–856. [PubMed: 25557322]
- Dunn, LM., Dunn, LM. Peabody Picture Vocabulary Test – Revised. Circle Pines, MN: American Guidance Service; 1981.
- Eigsti IM, Bennetto L, Dadlani MB. Beyond pragmatics: Morphosyntactic development in autism. *Journal of Autism and Developmental Disorders*. 2007; 37(6):1007–1023. [PubMed: 17089196]
- Feldman HM, Dale PS, Campbell TF, Colborn KD, Kurs-Lasky M, Rockette HE, Paradise JL. Concurrent and predictive validity of parent reports of child language at ages 2 and 3 years. *Child Development*. 2005; 76(4):856–868. [PubMed: 16026501]
- Fenson, L., Dale, PS., Reznick, JS., Thal, D., Bates, E., Hartung, J., Pethick, S., Reilly, J. The MacArthur Communicative Development Inventories: User's guide and technical manual. San Diego, CA: Singular Publishing Group; 1993.
- Geiger DM, Smith DT, Creaghead NA. Parent and professional agreement of cognitive level of children with autism. *Journal of Autism and Developmental Disorders*. 2002; 32(4):307–312. [PubMed: 12199135]
- Glaun DE, Cole KE, Reddihough DS. Mother – professional agreement about developmental delay in preschool children: A preliminary report. *Journal of Applied Research in Intellectual Disabilities*. 1999; 12(1):69–76.
- Johnson S, Wolke D, Marlow N. Developmental assessment of preterm infants at 2 years: Validity of parent reports. *Developmental Medicine & Child Neurology*. 2008; 50(1):58–62. [PubMed: 18173632]
- Just MA, Carpenter PA. The relation between comprehending and remembering some complex sentences. *Memory & Cognition*. 1976; 4(3):318–322. [PubMed: 21287042]
- Kelty-Stephen, E., Tek, S., Fein, D., Naigles, LR. Specific effects of joint attention on language development in children with autism spectrum disorders. In: Orman, W., Valteau, MJ., editors. *Proceedings of the 38th Annual Boston University Conference on Language Development*; Somerville, MA: Cascadilla Press; 2014. p. 205-215.

- Koegel LK, Koegel RL, Smith A. Variables related to differences in standardized test outcomes for children with autism. *Journal of Autism and Developmental Disorders*. 1997; 27:233–243. [PubMed: 9229256]
- Kover ST, Davidson MM, Sindberg HA, Weismer SE. Use of the ADOS for assessing spontaneous expressive language in young children with ASD: A comparison of sampling context. *Journal of Speech, Language, and Hearing Research*. 2014; 57(6):2221–2233.
- Kover ST, McCary LM, Ingram AM, Hatton DD, Roberts JE. Language development in infants and toddlers with Fragile X syndrome: Change over time and the role of attention. *American Journal on Intellectual and Developmental Disabilities*. 2015; 120(2):125–144. [PubMed: 25715182]
- Libertus ME, Odic D, Feigenson L, Halberda J. A Developmental Vocabulary Assessment for Parents (DVAP): Validating parental report of vocabulary size in 2–7-year-olds. *Journal of Cognition and Development*. 2015; 16(3):442–454.
- Lord C, Rutter M, Goode S, Heemsbergen J, Jordan H, Mawhood L, Schopler E. Autism diagnostic observation schedule: A standardized observation of communicative and social behavior. *Journal of Autism and Developmental Disorders*. 1989; 19(2):185–212. [PubMed: 2745388]
- Luyster RJ, Kadlec MB, Carter A, Tager-Flusberg H. Language assessment and development in toddlers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*. 2008; 38(8):1426–1438. [PubMed: 18188685]
- Massa J, Gomes H, Tartter V, Wolfson V, Halperin JM. Concordance rates between parent and teacher clinical evaluation of language fundamentals observational rating scale. *International Journal of Language and Communication Disorders*. 2008; 43(1):99–110. [PubMed: 17852530]
- McCarthy, D. McCarthy Scales of Children's Abilities. New York: The Psychological Corporation; 1972.
- MacWhinney, B. The Childe Project: Tools for Analyzing Talk. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc; 1995.
- Mervis CB, Klein-Tasman BP. Methodological issues in group-matching designs: α levels for control variable comparisons and measurement characteristics of control and target variables. *Journal of Autism and Developmental Disorders*. 2004; 34(1):7–17. [PubMed: 15098952]
- Mullen, E. Mullen Scales of Early Learning. Circle Pines, MN: American Guidance Service, Inc; 1995.
- Naigles, L., Chin, I. Language in children with autism spectrum disorders. In: Bavin, EL., Naigles, L., editors. *Cambridge Handbook of Child Language*. 2. Cambridge: Cambridge University Press; 2015. p. 637–658.
- Naigles LR, Cheng M, Rattanasone NX, Tek S, Khetrapal N, Fein D, Demuth K. "You're telling me!" The prevalence and predictors of pronoun reversals in children with autism spectrum disorders and typical development. *Research in Autism Spectrum Disorders*. 2016; 27:11–20. [PubMed: 27103941]
- Naigles, L., Fein, D. Looking through their eyes: Tracking early language comprehension in ASD. In: Naigles, L., editor. *Innovative Investigations of Language in Autism*. NY: APA Books/Walter deGruyter; 2017.
- Neale, MD. Neale Analysis of Reading Ability – Third Edition, Australian Edition. Melbourne, Victoria, Australia: Australian Council for Educational Research; 2007.
- Norbury, CF. The Cambridge Handbook of Communication Disorders. Cambridge: Cambridge University Press; 2013. Autism Spectrum Disorder; p. 141–158.
- Oliver B, Dale P, Saudino K, Petrill S, Pike A, Plomin R. The validity of a parent-based assessment of cognitive abilities in three-year olds. *Early Child Development and Care*. 2002; 172(4):337–348.
- Ozonoff S, Iosif AM, Young GS, Hepburn S, Thompson M, Colombi C, Cook IC, Werner E, Goldring S, Baquio F, Rogers SJ. Onset patterns in autism: Correspondence between home video and parent report. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2011; 50(8):796–806. [PubMed: 21784299]
- Park CJ, Yelland GW, Taffe JR, Gray KM. Morphological and syntactic skills in language samples of preschool aged children with autism: Atypical development? *International Journal of Speech-Language Pathology*. 2012; 14(2):95–108. [PubMed: 22390743]

- Ring ED, Fenson L. The correspondence between parent report and child performance for receptive and expressive vocabulary beyond infancy. *First Language*. 2000; 20(59):141–159.
- Semel, E., Wiig, EH., Secord, WA. *Clinical Evaluation of Language Fundamentals*. 3. San Antonio, TX: Psychological Corporation; 1995.
- Sparrow, SS., Cicchetti, DV., Balla, DA. *Vineland Adaptive Behavior Scales – Second Edition*. Circle Pines, MN: AGS; 2005.
- Stone WL, Coonrod EE, Ousley OP. Brief report: Screening tool for autism in two-year-olds (STAT): Development and preliminary data. *Journal of Autism and Developmental Disorders*. 2000; 30(6): 607–612. [PubMed: 11261472]
- Styles S, Plunkett K. What is ‘word understanding’ for the parent of an one-year-old? Matching the difficulty of a lexical comprehension task to parental CDI report. *Journal of Child Language*. 2009; 36(4):895–908. [PubMed: 19079845]
- Tager-Flusberg, H., Paul, R., Lord, CE. Language and communication in autism. In: Volkmar, F.Paul, R.Klin, A., Cohen, DJ., editors. *Handbook of autism and pervasive developmental disorder*. 3. Vol. 1. New York: Wiley; 2005. p. 335-364.
- Tek S, Mesite L, Fein D, Naigles L. Longitudinal analyses of expressive language development reveal two distinct language profiles among young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*. 2014; 44(1):75–89. [PubMed: 23719855]
- Wilkinson, GS., Robertson, GJ. *Wide Range Achievement Test*. 4. Lutz, FL: Psychological Assessment; 2006.

Highlights

- Seven grammatical items are compared with children's usage in their spontaneous speech on the same day.
- Parents of Middle Verbal-ASD children confirmed relatively few of their children's productions of the seven grammatical items, even though the children produced them.
- Pronouns were the most difficult for parents to catch in all three groups.
- Past tense verbs were also frequently missed by mothers of High Verbal-ASD children.

What this paper adds

This paper investigates the level of agreement between parent report and child speech in TD children and children with ASD. Previous research has found good reliability between experimenter assessments and parent reports; however, earlier studies have not examined parent report of children's use of *specific* grammatical items. The present study conducted a detailed comparison of children's usage of grammatical constructions in their spontaneous speech and as reported by parent report on the *same* day. The results found that parents of TD and high-verbal-ASD children confirmed children's productions at high rates at both visits whereas parents of Middle-verbal children with ASD confirmed fewer of their child's productions at both early and later visits. With respect to individual grammatical items, pronouns were the most challenging, as mothers frequently reported 'does not say' for these when their children did produce pronouns at the same visit. Parents of High-verbal children with ASD also frequently missed when their children produced past tense verbs. Our findings indicate that this level of inaccuracy puts limits on interpreting parent reports of children's emerging language.

Figure 1.

Maternal Percent Inaccurately Overlooked by Each Item by Group and Visit.

□ Represents average level for each group; ■ Represents items higher than average; ▨ Represents items lower than average.

Table 1

Demographic and Standardized Test Data from Typically Developing (TD) and Autism Spectrum Disorder (ASD) Groups at First and Final Visits.

	TD	HV-ASD	MV-ASD
Visit 1			
<i>N</i>	24	9	11
	Mean (SD)	Mean (SD)	Mean (SD)
Age in months	20.40 (1.67)	31.34 (6.04)	32.77 (5.67)
Range	18.80 – 24.40	18.77 – 39.50	24.93 – 42.00
CDI – Number of words produced	104.46 (104.24)	224.33 (105.05)	42.54(39.74)
Range	11– 317	10 – 335	0–96
ADOS *	0.46 (0.88)	10.11 (2.71)	14.00 (2.76)
Range	0–3	7–14	10–18
Mullen T-Scores ⁺			
Visual Reception	57.96 (10.05) ^a	53.00 (10.09) ^a	34.91 (12.22) ^b
Fine Motor	48.58 (7.57) ^a	47.33 (16.14) ^a	28.09 (1.86) ^b
Receptive Language	56.79 (11.98) ^a	60.44 (12.23) ^a	31.09 (10.42) ^b
Expressive Language	47.67 (12.56) ^a	45.33 (10.09) ^a	27.55 (8.82) ^b
Mullen Raw Scores ⁺			
Visual Reception	26.08 (3.28) ^a	32.78 (4.27) ^b	26.46 (3.14) ^a
Fine Motor	21.88 (1.70) ^a	29.11 (3.10) ^b	23.55 (1.86) ^b
Receptive Language	24.25 (3.59) ^a	32.89 (4.17) ^b	20.64 (5.37) ^a
Expressive Language	19.29 (4.91) ^a	26.56 (5.17) ^b	16.82 (4.21) ^a
Vineland Standard Scores ⁺			
Communication	102.63 (7.72) ^a	95.22 (11.87) ^a	73.36 (12.33) ^b
Daily Living	105.21 (8.50) ^a	88.11 (17.54) ^b	75.09 (9.27) ^b
Socialization	100.46 (6.89) ^a	78.78 (8.93) ^b	74.36 (6.27) ^b
Motor	101.75 (5.69) ^a	92.00 (12.31) ^b	84.18 (12.14) ^b
Visit 6			
Mullen T- Scores ⁺			
Visual Reception	64.92 (9.98) ^a	59.11 (11.48) ^a	37.73 (17.36) ^b
Fine Motor	51.71 (10.64) ^a	52.33 (18.54) ^a	24.73 (6.74) ^b
Receptive Language	59.42 (10.11) ^a	59.89 (13.21) ^a	30.82 (12.82) ^b
Expressive Language	58.67 (10.28) ^a	58.89 (14.43) ^a	25.55 (8.27) ^b
Mullen Raw Scores ⁺			
Visual Reception	43.88 (3.37) ^a	46.67 (3.64) ^b	39.18 (7.14) ^a
Fine Motor	36.46 (3.28) ^a	42.11 (3.92) ^b	32.64 (4.82) ^b
Receptive Language	39.08 (4.13) ^a	44.11 (4.43) ^b	32.18 (5.64) ^b
Expressive Language	38.96 (5.15) ^a	44.67 (4.97) ^b	29.18 (4.87) ^b
Vineland Standard Scores ⁺			

	TD	HV-ASD	MV-ASD
Communication	104.67 (9.48) ^a	104.78 (7.17) ^a	87.27 (12.94) ^b
Daily Living	103.04 (9.09) ^a	95.00 (21.34) ^a	79.18 (14.15) ^b
Socialization	99.71 (6.99) ^a	89.44 (14.26) ^a	77.27 (14.59) ^b
Motor	103.67 (9.72) ^a	101.56 (15.37) ^a	85.46 (10.51) ^b

* Cut-off score for a diagnosis of autism is 12, and cut-off score for a diagnosis of autism spectrum disorders is 7.

⁺ Different superscripts, e.g., ^a and ^b signify that the groups (i.e. TD vs. HV-ASD or TD vs. MV- ASD) are significantly different on the standardized subdomain. The same superscripts, e.g., ^a and ^a represent that the groups (e.g. TD vs. HV-ASD or TD vs. MV-ASD) did not significantly differ on the particular subdomain.

Table 2

Mean Percentage (SD) of Transcripts Demonstrating Child Production of Vineland Items by Group and Visit.

<u>Visit/Measure</u>	<u>Group</u>		
	TD	HV-ASD	MV-ASD
Number of Children	24	9	11
Early visit (1–3) transcripts	72	27	32
% of Transcripts that include Child Productions, averaged across 7 items	44.05% (18.32) ^a	70.90% (21.00) ^b	20.78% (14.26) ^c
Later visit (4–6) transcripts	72	26	32
% of Transcripts that include Child Productions, averaged across 7 items	85.12% (20.97) ^a	91.53% (17.98) ^a	45.02% (9.05) ^b

Note. Different superscripts, e.g., ^a and ^b signify that the groups (i.e. TD vs. HV-ASD, TD vs. MV- ASD and HV-ASD vs. MV-ASD) are significantly different.

Table 3

Mean Percentage (SD) of Maternal Reports of ‘Doesn’t say’, by Group and Visit.

<u>Visit/Measure</u>	<u>Group</u>		
	TD	HV-ASD	MV-ASD
Number of Children	24	9	11
Early visit (1–3) administrations	50	23	15
% Reports of ‘doesn’t say’ ¹	37.15% (10.62) ^a	28.82% (28.69) ^a	60.83% (14.73) ^b
Later visit (4–6) administrations	72	26	26
% Reports of ‘doesn’t say’ ²	6.51% (2.13)	11.11% (5.24)	36.96% (23.59)

Note. Different superscripts, e.g., ^a and ^b signify that the groups (i.e. TD vs. HV-ASD, TD vs. MV- ASD and HV-ASD vs. MV-ASD) are significantly different.

¹ averaged across 7 items for TD group, 6 items for HV-ASD group, and 7 items for MV-ASD group.

² averaged across 4 items for TD group, 2 items for HV-ASD group, and 7 items for MV-ASD group.

Table 4

Mean Percentage (SD) of Maternal Accurate Confirmed Productions, by Group and Visit.

<u>Visit/Measure</u>	<u>Group</u>		
	TD	HV-ASD	MV-ASD
Early visit (1–3) transcripts	72	27	32
% Accurate Confirmations	79.50% (14.33) ^a	78.80% (23.56) ^a	57.52% (33.25) ^a
Later visit (4–6) transcripts	72	26	32
% Accurate Confirmations	96.18% (4.85) ^a	97.16% (5.07) ^a	64.07% (33.68) ^b

Note. Different superscripts, e.g., ^a and ^b signify that the groups (i.e. TD vs. HV-ASD, TD vs. MV- ASD and HV-ASD vs. MV-ASD) are significantly different.

Table 5

Mean Percentage (SD) of Maternal Inaccurately Overlooked Structures, by Group and Visit.

<u>Visit/Measure</u>	<u>Group</u>		
	TD	HV-ASD	MV-ASD
Early visit (1–3) administrations	50	23	15
% Inaccurately Overlooked Maternal Reports ¹	35.57% (31.37) ^a	43.39% (43.04) ^a	32.76% (31.65) ^a
Later visit (4–6) administrations	72	26	26
% Inaccurately Overlooked Maternal Reports ²	65.83% (35.53)	87.5% (17.68)	51.0% (37.42)

Note. Different superscripts, e.g., ^a and ^b signify that the groups (i.e. TD vs. HV-ASD, TD vs. MV- ASD and HV-ASD vs. MV-ASD) are significantly different.

¹ averaged across 7 items for TD group, 6 items for HV-ASD group, and 7 items for MV-ASD group.

² averaged across 4 items for TD group, 2 items for HV-ASD group, and 7 items for MV-ASD group.